BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) Europäisch s Patentamt
European Patent Office
Offic uropé n des brev ts

(11) EP 1 022 724 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
26.07.2000 Bulletin 2000/30

(51) Int Cl.7: **G10L 15/06**

(21) Application number: **99100952.3**

(22) Date of filing: **20.01.1999**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **Sony International (Europe) GmbH**
10785 Berlin (DE)

(72) Inventors:
- **Gorozny, Silke c/o Sony Int. (Europe) Gmbh**
70736 Fellbach (DE)
- **Kompe, Ralf c/o Sony Int. (Europe) Gmbh**
70736 Fellbach (DE)
- **Buchner, Peter c/o Sony Int. (Europe) Gmbh**
70736 Fellbach (DE)

(74) Representative:
MÜLLER & HOFFMANN Patentanwälte
Innere Wiener Strasse 17
81667 München (DE)

(54) Speaker adaptation for confusable words

(57) A well-known problem in state-of-the-art speech recognition system is that often pairs of words occur that are very similar to each other and thus are confusable. This may cause errors in the recognition phase and thus decrease recognition rates. If unsupervised speaker adaptation is used in such a system, these misrecognitions may cause adaptation of the wrong models and thus cause a further decrease in performance. Therefore, according to the present invention, confusable words within the vocabulary are marked and an adaptation of the system to a certain user with such marked words is only performed in case of a positive confirmation of the recognition result by the user.

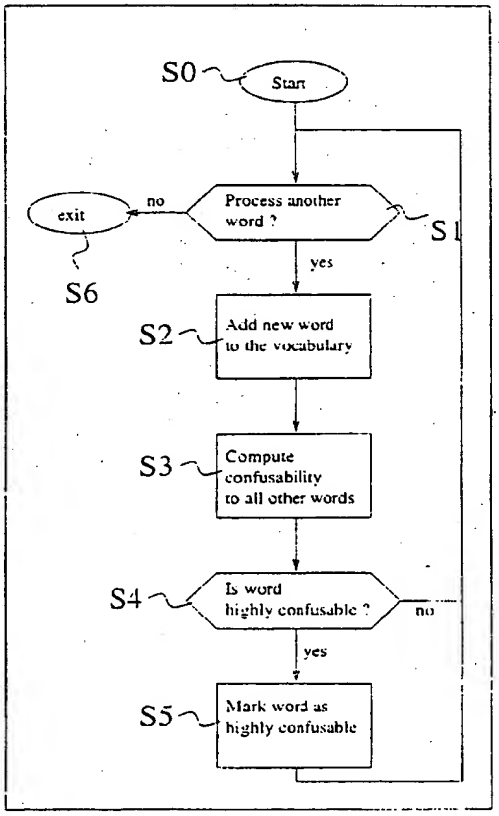

Fig. 1

EP 1 022 724 A1

Description

[0001] This invention is related to a method to perform an adaptation of an automatic speech recognition system, in particular to a method to prevent an adaptation of the wrong models in a speech recognition system.

[0002] State of the art speech recognizers consist of a set of statistical distribustions modeling the acoustic properties of certain speech segments. These acoustic properties are encoded in feature vectors. As an example, one Gaussian distribution can be taken for each phoneme. These distributions are attached to states. A (stochastic) state transition network (usually hidden Markov models) defines the probabilities for sequences of states and sequences of feature vectors. Passing a state consumes one feature vector covering a frame of e.g. 10 ms of the speech signal.

[0003] The stochastic parameters of such a recognizer are trained using a large amount of speech data either from a single speaker yielding a speaker dependent (SD) system or from many speakers yielding a speaker independent (SI) system.

[0004] Speaker adaptation (SA) is a widely used method to increase recognition rates of SI systems. State of the art speaker dependent systems yield much higher recognition rates than speaker independent systems. However, for many applications, it is not feasible to gather enough data from a single speaker to train the system. In case of a consumer device this might even not be wanted. To overcome this mismatch in recognition rates, speaker adaptation algorithms are widely used in order to achieve recognition rates that come close to speaker dependent systems, but only use a fraction of speaker dependent data compared to speaker dependent ones. These systems initially take speaker independent models that are then adapted so as to better match the speaker's acoustics.

[0005] Usually, the speaker adaptation is performed in supervised mode. That is the spoken words are known and the recognizer is forced to recognize them. Herewith a time alignment of the segment-specific distributions is achieved. The mismatch between the actual feature vectors and the parameters of the corresponding distribution builds the basis for the adaptation. The supervised adaptation requires an adaptation session to be done with every new speaker before he/she can actually use the recognizer.

[0006] Usually, the speaker adaptation techniques modify the parameters of the hidden Markov models so that they better match the new speakers acoustics. Normally, in batch or off-line adaptation a speaker has to read a pre-defined text before he/she can use the system for recognition, which is then processed to do the adaptation. Once this is finished the system can be used for recognition. This mode is also called supervised adaptation, since the text was known to the system and a forced alignment of the corresponding speech signal to the models corresponding to the text is performed and used for adaptation.

[0007] However, an unsupervised or on-line method is better suited for most kinds of consumer devices. In this case, adaptation takes place while the system is in use. The recognized utterance is used for adaptation and the modified models are used for recognizing the next utterance and so on. In this case the spoken text is not known to the system, but the word(s) that were recognized are taken instead.

[0008] An adapatation of the speaker adapted model set can be repreatedly performed to further improve the performance of specific speakers. There are several existing methods for speaker adaptation, e.g. maximum a posteriori adaptation (MAP) or maximum likelihood linear regression (MLLR) adaptation.

[0009] For speech recognition systems often the problem arises that the vocabulary comprises many words that sound similar. As a consequence it is often difficult to distinguish between these words and this often causes misrecognitions. If a system uses unsupervised speaker adaptation to improve its models for particular speakers, these misrecognitions may lead adaptation to the wrong direction and this may have an adverse effect on the recognition rates, since then the wrong models are modified.

[0010] State-of-the-art speech recognition systems try to resolve ambiguities using grammars and language models that define a structure of valid sentences, so that in some cases ambiguities can be resolved by this.

[0011] Another method disclosed in EP 0 763 812 A1 is the use of verification methods to reduce the confusability of certain words. It is a mathematical approach in which confidence measures are used for verification of n-best recognized words strings. The result of this verification procedure (the derivative of the loss function) is used as an optimization criterion for HMM training prior to the use of the system. In this case, all utterances are used for training and the method is used to maximize the difference in the likelihood of confusable words.

[0012] However, in supervised or especially in unsupervised speech recognition systems misrecognitions can occur so that then the wrong HMMs will be adapted. If this happens repeatedly, recognition performance may decrease drastically.

[0013] Therefore, it is the object underlying the present invention to propose a method for adaptation that overcomes the problems described above.

[0014] The inventive method is defined in independent claim 1. Preferred embodiments thereof are defined in the respective following dependent claims.

[0015] This problem is solved by avoiding adaptation based on a misrecognized word if this is confusable, e. g. highly confusable with other words.

[0016] According to the inventive method the speech recognition system is made aware of such highly confusable words and if such a word is recognized, double checks the recognition result by asking for confirmation from the user. Only when the system can be sure that

such a word was recognized correctly, it will be used for adaptation.

[0017] Therefore, prior to the recognition phase, it is determined which words in the vocabulary are highly confusable with other words. This is e.g. done by comparing and computing the number of differing phonemes in relation to the total number of the words. Another possibility might be to use sets of template speech signals representing all words in the vocabulary and then computing a distance between these words. Such templates can preferably be Hidden Markov Models. Of course the determination of the confusability is not limited to this.

[0018] Also, a grade of confusabiltity of a certain word contained in the vocabulary to the other words of the vocabulary can be determined. This can be done manually or automatically using well known similarity measures for phoneme strings and/or HMMs. In this case not only highly confusable words, but also words that are confusable at all have to get a confirmation or may also or instead be processed with other verification technologies.

[0019] In any case, for each word in the vocabulary it is known if and with which word(s) it is confusable and the grade of this respective confusability. If during recognition one of the words that was previously marked as being confusable is recognized, e.g. as highly confusable, the user is asked to confirm the recognition result and in case it was misrecognized, to repeat or spell it (if the user interface comprises a keyboard he/she could also type it; other input modalities are also suitable for correction purposes). After that is done the system can use the speech signal of the previously misrecognized words for which it now knows the correct word for adaptation. If the word was not a confusable one, no confirmation from the user is needed but other methods to verify the reliability of the recognition results may be applied.

[0020] As a result of the inventive method the confusability of generally confusable words will decline, because always the right models are adapted and thus the discrimination for highly confusable words should become easier.

[0021] The inventive method to perform an adaptation of an automatic speech recognition system will be better understood from the following detailed description of an exemplary embodiment thereof taken in conjunction with the appendent drawings wherein:

Figure 1 shows the process of determination of confusability between words in the vocabulary according to the present invention; and

Figure 2 shows the procedure to perform an adaptation according to the present invention.

[0022] Figure 1 shows the determination of confusability between words in the vocabulary prior to the recognition phase. According to the exemplary embodiment, it is determined here which words in the vocabulary are highly confusable with other words.

[0023] After the start of the procedure in a step S0, it is confirmed in a step S1 whether to process another word or not. In case no additional word should be processed the procedure will be set forth with step S6 to be ended. In the other case the procedure will be set forth with step S2 in which a new word is added to the vocabulary. Thereafter, the confusability of this new word to all other words already contained in the vocabulary is computed in a step S3.

[0024] As mentioned above, there are several methods to compute the confusability and also several results, i.e. several grades of confusability.

[0025] In a step S4 it is determined if this new word is highly confusable to other words. However, the determination whether this word is highly confusable with other words is no limitation of the present invention. It could also be checked in this step S4 if the word is confusable with other words at all and the grade of confusability will be passed on to the next step. In this exemplary embodiment, however, a classification of the grade of confusability is already performed in step S4.

[0026] If the new word is highly confusable with other words, this word is marked as highly confusable in a step S5. Thereafter, the procedure is set forth again with step SI in which it is checked if another word should be processed. If the word is not regarded as to be highly confusable with other words in step S4, the procedure is directly set forth with step S1.

[0027] Figure 2 shows the endless loop in which an adaptation according to the present invention is performed during the recognition process.

[0028] After an utterance of a user was spoken to the system in a step S9, a recognition of this utterance is performed in a step S10. In a step S11 it is checked whether one of the recognized words is highly confusable with other words (that are possible in this context) or not.

[0029] If the recognized word is regarded as to be highly confusable with at least one other word of the vocabulary in step S11, the user is asked for confirmation of the word in step S12. After the user's confirmation an adaptation of the models is performed in step S13 and the next spoken utterance is received in step S9.

[0030] If the recognized word is not regarded as to be highly confusable with any other words in step S11 any other verification method can be applied to the word in step S14. Thereafter, it is checked whether the word was recognized correctly or not in step S15. If the word was recognized correctly in step S15, an adaptation is performed in step S13, whereafter the next spoken utterance is received in step S9. If the word was not recognized correctly the next spoken utterance is directly received in step S9 without performing the adaptation in step S13.

Claims

1. Method to perform an adaptation of an automatic speech recognition system, **charact riz d by** the following steps:

   marking the words in the vocabulary of the speech recognition system that are confusable with other words prior to the recognition process;
   requesting a confirmation during the recognition process when a word marked as confusable is recognized; and
   perform an adaptation of the automatic speech recognition system with recognized words marked as confusable for which a positive confirmation was given.

2. Method according to claim 1, **characterized by** the following step:

   verify recognized words not marked as confusable; and
   perform an adaptation of the automatic speech recognition system with verified words.

3. Method according to claim 1 or 2, **characterized by** the following step:
   perform an adaptation of the automatic speech recognition system with recognized words not marked as confusable.

4. Method according to anyone of claims 1 to 3, **characterized in that** a word in the vocabulary of the speech recognition system is marked as confusable on the basis of a comparison and computation of the number of differing phonemes in relation to the total number of phonemes of the word.

5. Method according to anyone of claims 1 to 4, **characterized in that** a word in the vocabulary of the speech recognition system is marked as confusable on the basis of a computation of a distance measure between all words in the vocabulary with the help of a set of template speech signals representing all words in the vocabulary.

6. Method according to claim 5, **characterized in that** said set of template speech signals are Hidden Markov Models.

7. Method according to anyone of claims 1 to 6, **characterized in that** in case of a negative confirmation the user is asked to repeat the misrecognized word.

8. Method according to anyone of claims 1 to 7, **charact rized in that** in case of a negative confirmation the user is asked to spell the misrecognized word.

9. Method according to claim 7 or 8, **charact rized in that** the previously misrecognized word is used for adaptation after its correct recognition on basis of the repetition and/or spelling.

10. Method according to anyone of claims 1 to 9, **characterized in that** in case of a negative confirmation the user is asked to type in the misrecognized word.

11. Method according to anyone of claims 1 to 10, **characterized in that** the adaptation of the speech recognition system is an adaptation of the speaker independent Hidden Markov Models to speaker adapted Hidden Markov Models.

12. Method according to claim 11, **characterized in that** the adaptation method is maximum aposteriori adaptation or a maximum likelihood linear regression adaptation.

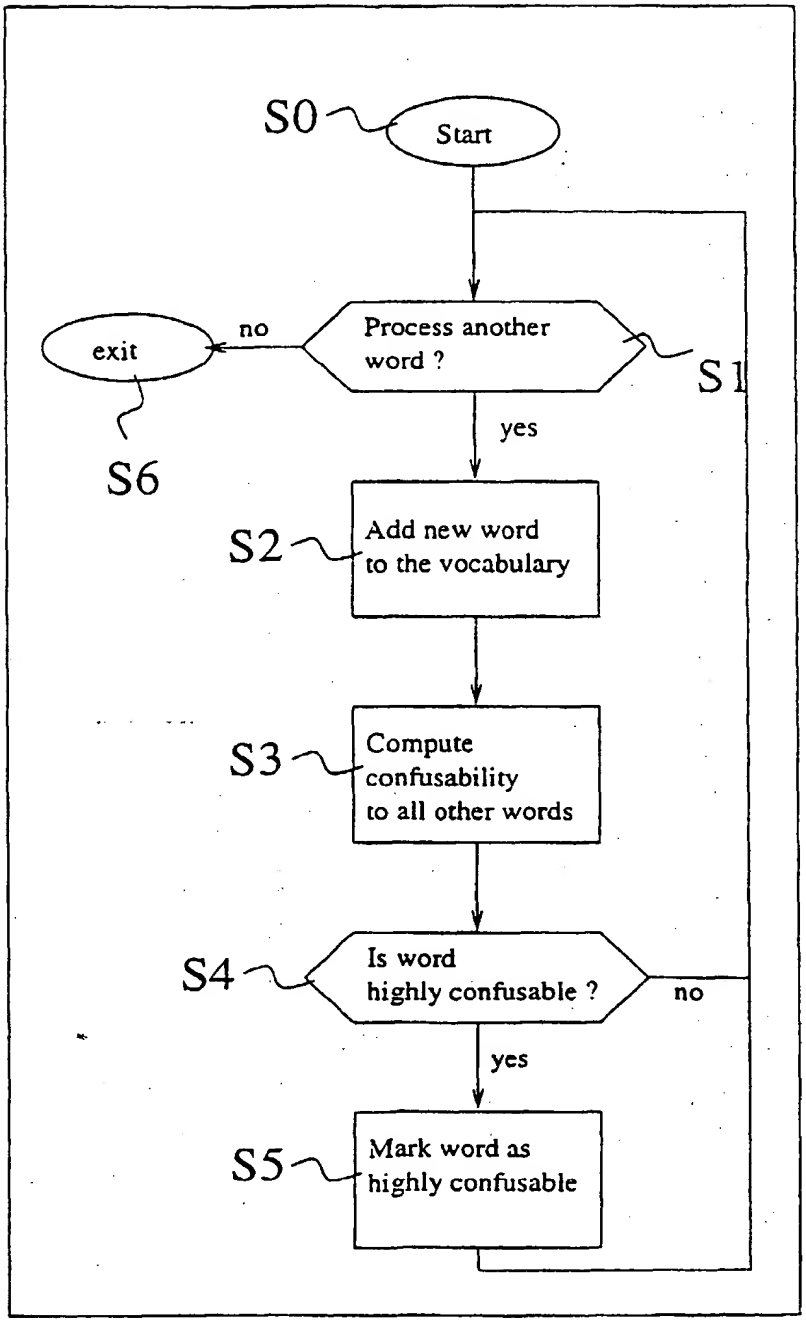

Fig. 1

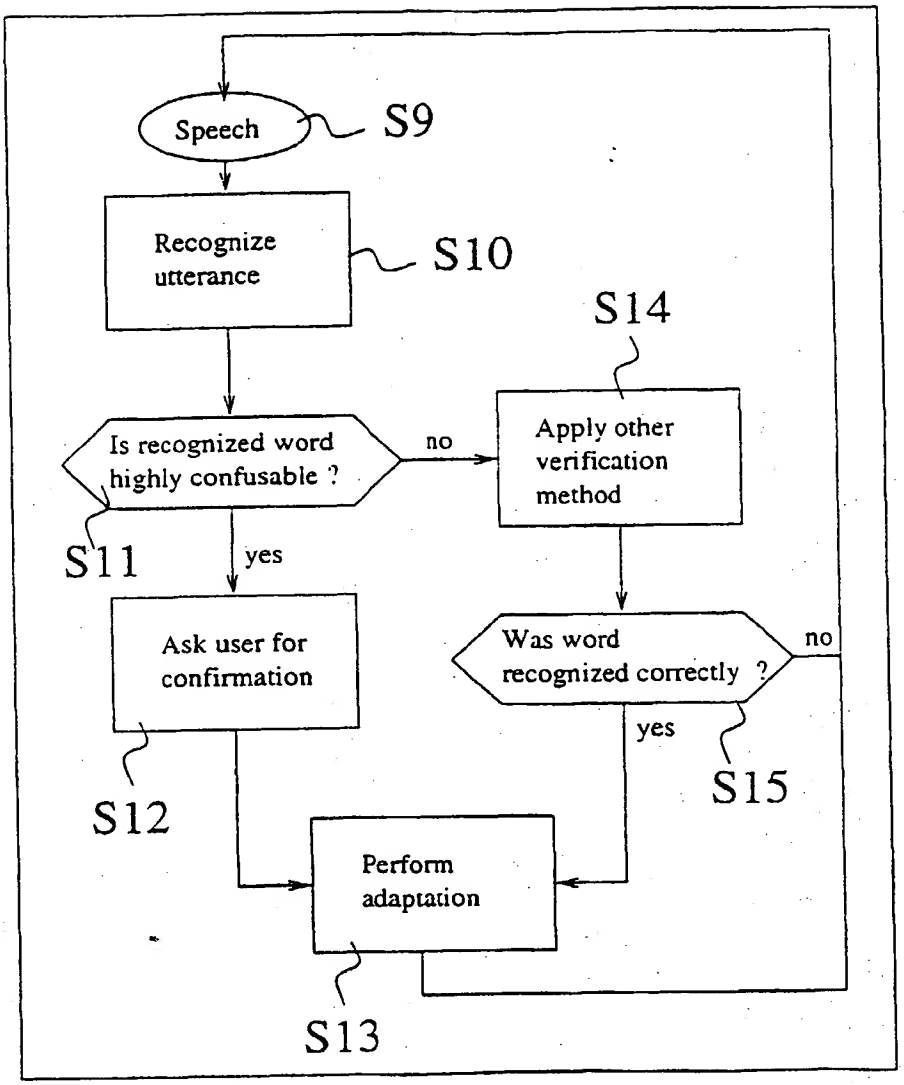

Fig. 2

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 99 10 0952

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION
Y	WO 95 28790 A (NORTHERN TELECOM) 26 October 1995 (1995-10-26) * page 6, line 14 - line 33 *	1	G10L15/06
Y	EP 0 700 031 A (AT & T) 6 March 1996 (1996-03-06) * page 2, line 36 - line 51 *	1	
A	EP 0 836 144 A (MICROSOFT) 15 April 1998 (1998-04-15) * column 1, line 56 - column 2, line 39 *	1	
A	ANONYMOUS: "Corrective Training of Feneme-Based Markov Models for Discrimination between Highly Confusable Words" IBM TECHNICAL DISCLOSURE BULLETIN, vol. 36, no. 4, April 1993 (1993-04), pages 533-536, XP000364607 New York, US * the whole document *	1	

TECHNICAL FIELDS SEARCHED
G10L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	7 July 1999	Lange, J

CATEGORY OF CITED DOCUMENTS

X particularly relevant if taken alone
Y particularly relevant if combined with another document of the same category
A technological background
O non-written disclosure
P intermediate document

T theory or principle underlying the invention
E earlier patent document, but published on, or after the filing date
D document cited in the application
L document cited for other reasons

& member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 99 10 0952

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

07-07-1999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9528790 A	26-10-1995	US 5488652 A	30-01-1996
		CA 2185356 A	26-10-1995
		EP 0757868 A	12-02-1997
		JP 9505710 T	03-06-1997
		US 5644680 A	01-07-1997
EP 0700031 A	06-03-1996	US 5737723 A	07-04-1998
		JP 8076788 A	22-03-1996
EP 0836144 A	15-04-1998	CN 1172992 A	11-02-1998
		JP 10078964 A	24-03-1998

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82